# Towards a Virtual Heliospheric Observatory: A Prototype Example of Distributed Data Sharing

T.W. Narock [1,2], A. Szabo[1]

1. Laboratory for Extraterrestrial Physics, NASA Goddard Space Flight Center, Greenbelt, MD. 20771
2. L3 Communications, GSI, Chantilly, VA. 20151

Tom.Narock@gsfc.nasa.gov

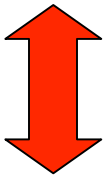
AGU 2003 Fall Meeting U22A-0016

# Abstract

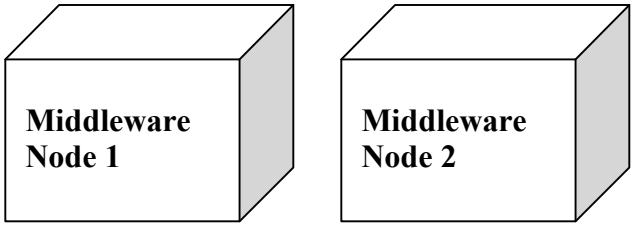
In the current space science paradigm various data sets are available from different data providers (often located in different countries), hence, necessitating the development of methods and standards that would allow for effortless search and retrieval of this vast distributed data. Such methods and services must ingest different data formats as well as standards. A prototype system has been developed using IMP8 and WIND magnetometer data sets and we aim to discuss its design, use and expansion to incorporate all near-Earth spacecraft. This all encompassing system will be known as the Virtual Heliospheric Observatory (VHO).

# Requirements

1.) Must provide the users with an easy to use interface by which they can query several data providers simultaneously

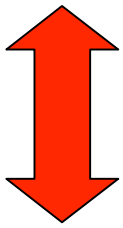


2.) The interface must be lightweight and easy to use

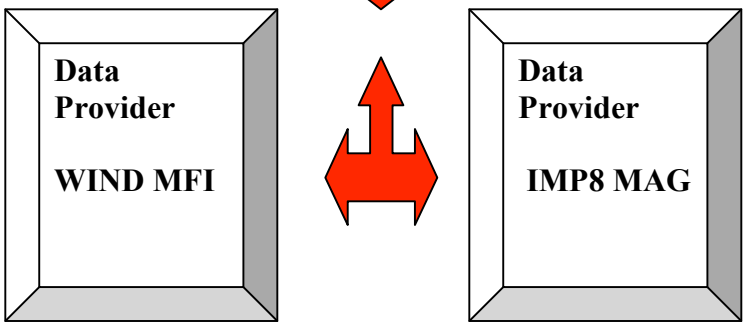



3.) Initial knowledge, by the user, of the data sets or how they are described should not be required. Handled by middleware.

4.) Options should be available for users to obtain available search parameters at the time of query

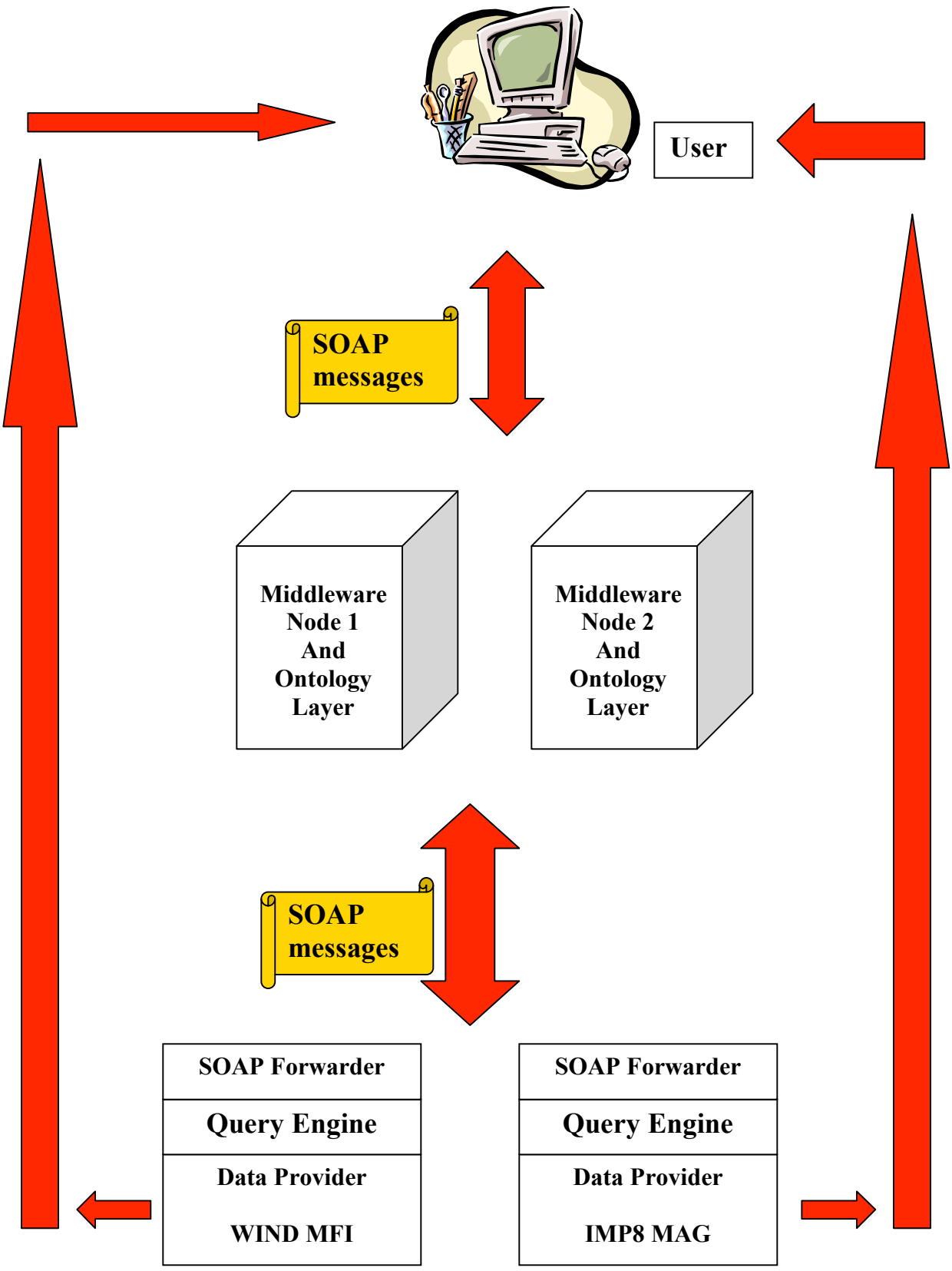


5.) The system must process queries in a reasonable amount of time

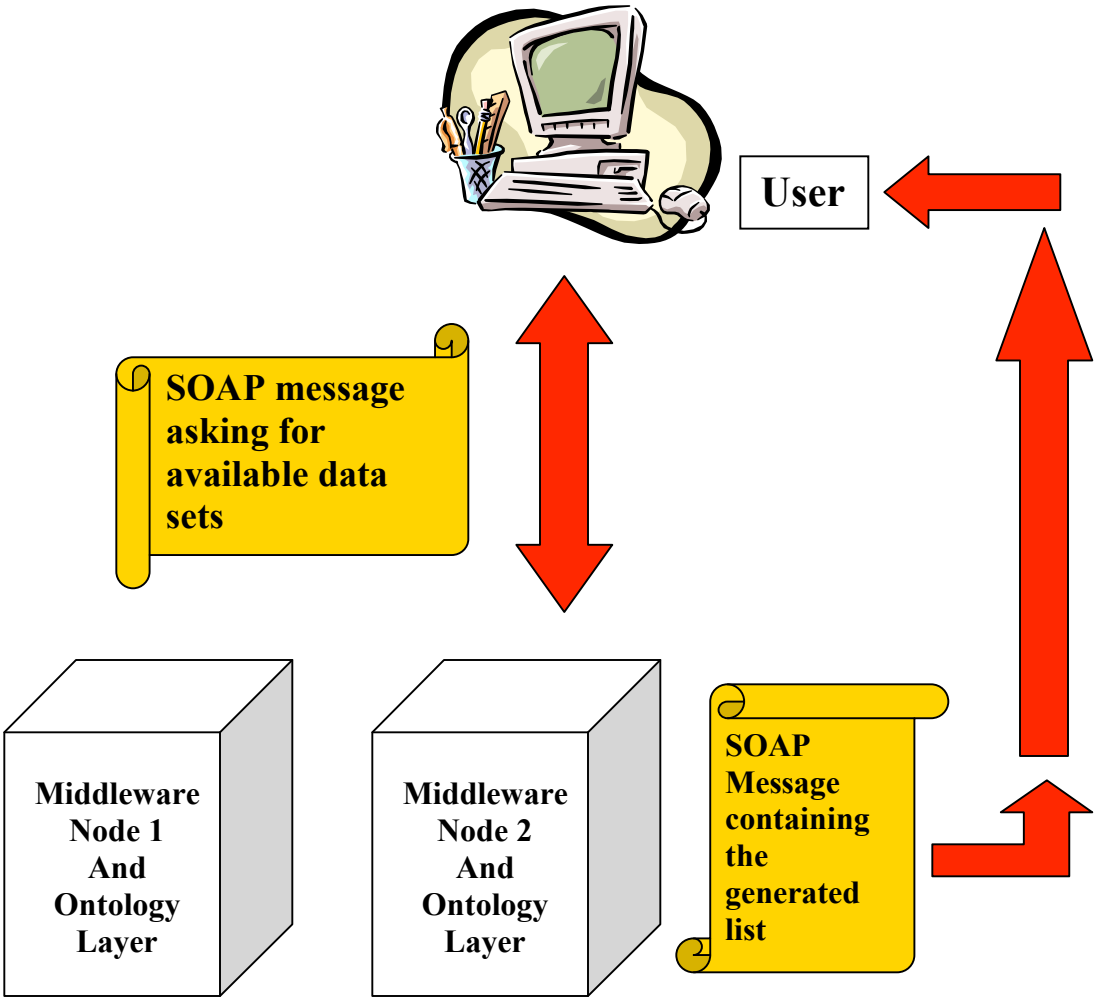



6.) Methods should be in place to retrieve the data once results have been obtained along with reading software so user does not have to learn various storage formats

# System Design



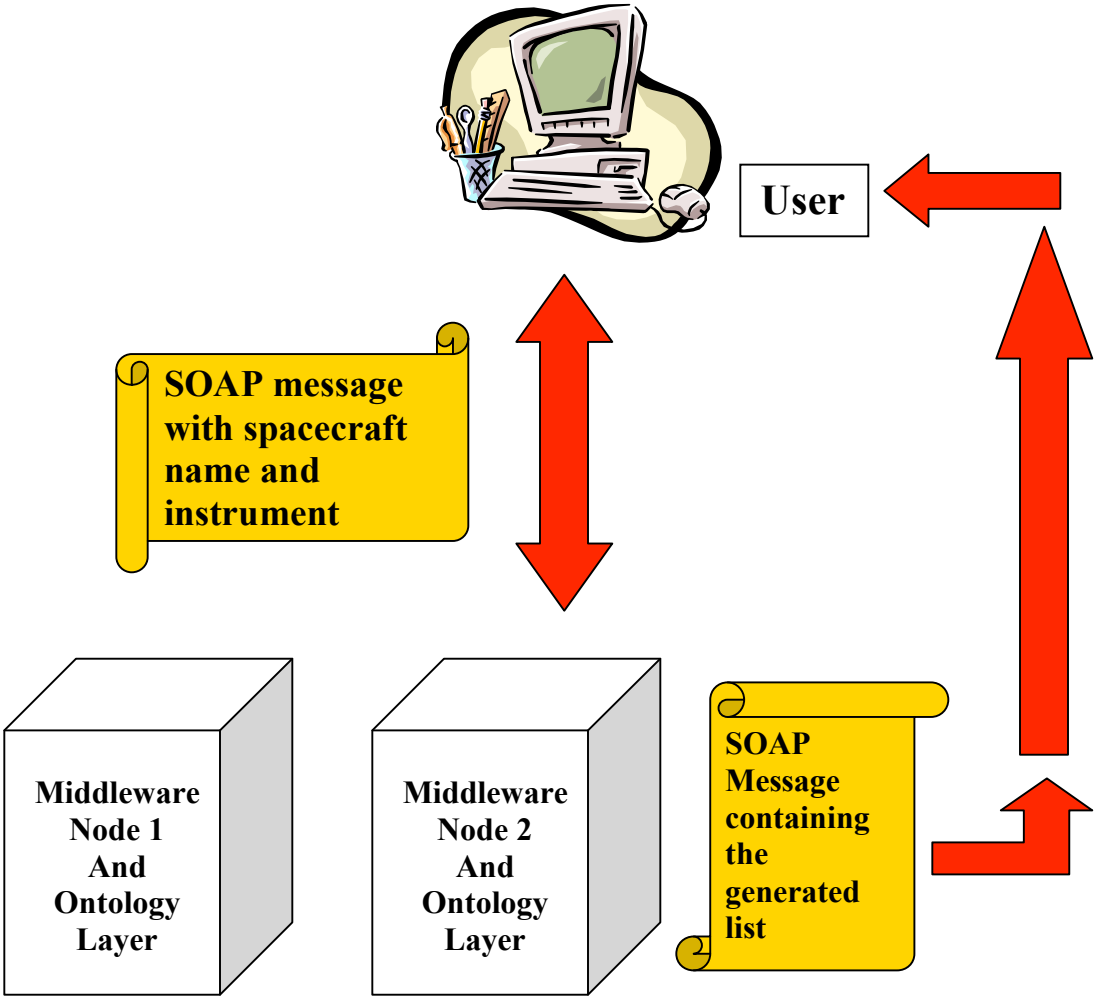
# Finding Available Data




Middleware Node contains RDF file for each data provider available to the system

A list is generated containing spacecraft name, instrument name and instrument type for all data providers

# Finding Available Search Parameters




**Middleware Node reads RDF data stored in memory and generates a listing of available search parameters based on the incoming spacecraft name and instrument name. Returned listing is generated using the SPASE+ dictionary so users see the same name for all data sets containing identical quantities.**

# Submitting a Query

User

SOAP message with queries and names of all data providers to whom they should be submitted

Middleware Node 1 And Ontology Layer

Middleware Node 2 And Ontology Layer

SOAP Message containing the number of files matching the query and error flags for each provider submitted too

Middleware Node receives the query and submits them in parallel to the data providers. A unique ID is obtained from each data provider and the middleware node checks for results using the ID. Either all results are obtained or the Time to Live expires. At this point all obtained results are returned to the user.

Data Provider(s)

# Downloading Data

User

SOAP message with ID number and instrument name

Middleware Node 1 And Ontology Layer

Middleware Node 2 And Ontology Layer

Data Transfer

Data Provider(s)

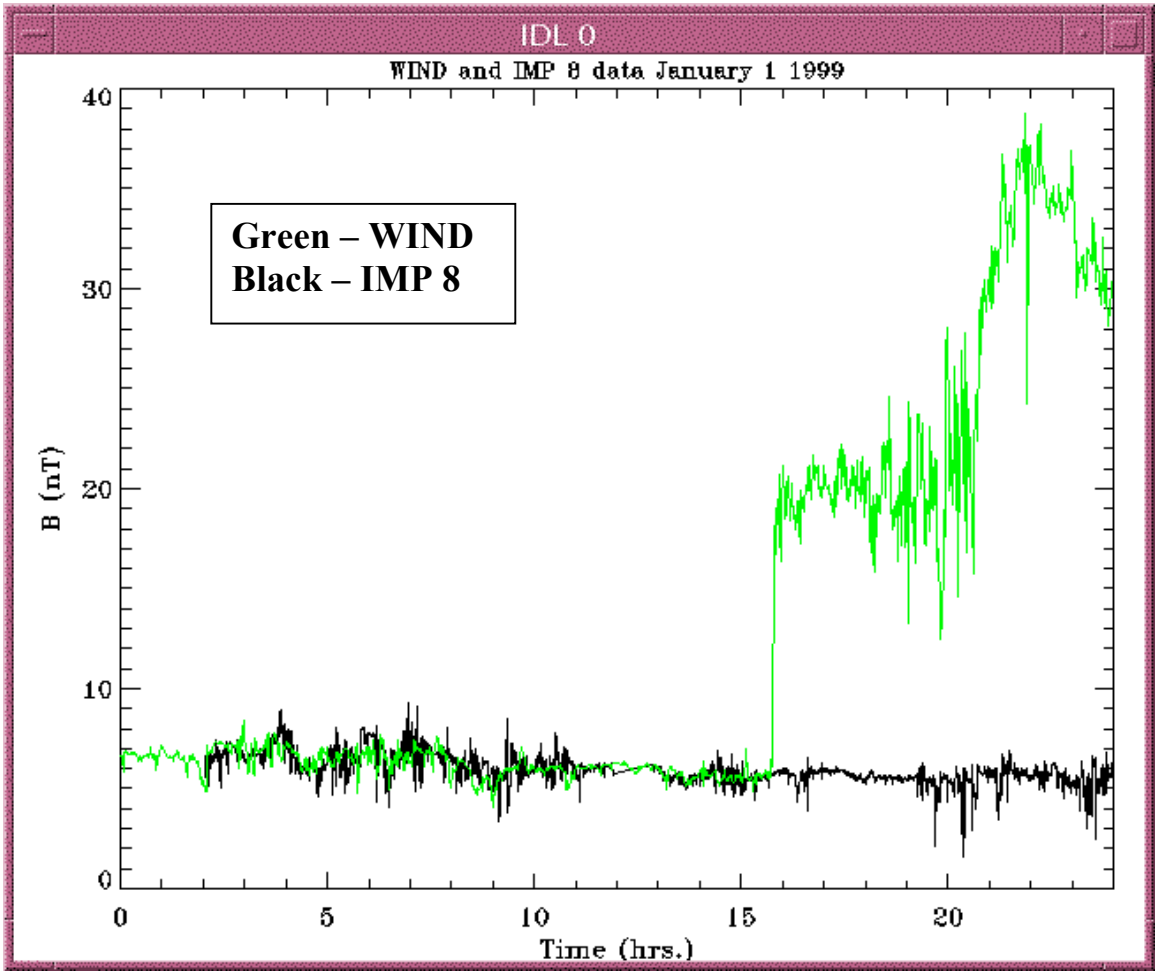
Middleware Node receives the request and places the user in contact with the data provider. A TCP/IP connection is established and user software uses ID number to obtain a file listing that was generated during query. Files are transferred one at a time in kilobyte chunks through TCP/IP connection.

# User Software

- Consists of both command line and graphical user interface implementations
- Users are free to write their own implementations or work current interfaces into their own software
- SOAP access supplied through Perl code. SOAP libraries are also available for a number of other programming languages including Java and C++
- Provides users with 4 options: finding what data sets are available, finding searchable parameters in each of these data sets, submission of queries and downloading of files that meet queries
- These 4 options can be carried out in any order and are not dependant upon each other

# Post-Download Software

- User should not have to become familiar with various data storage formats
- Software exists to read and visualize CDF, HDF and ascii data
- User needs one reading routine and global metadata file and the software handles the rest



# Security and Error Handling

Security

- Messaging between user and middleware as well as between middleware and data providers is done over SSL (assuming user has installed SSL, not required but strongly recommended)
- User never sees hostname, IP address or port numbers of data providers
- Users also never send messages directly to data providers only placed in contact with them to download data

SOAP Errors

- SOAP messages sent over HTTP protocol meaning any errors are returned as HTTP error codes
- Error description and handling can be kept to a minimum as it is only necessary to interpret and handle 400 and 500 series HTTP codes

# Added Value Services

User

A complete and comprehensive data environment needs post-processing of the data

1.) User software sends SOAP message to middleware asking for positions of a spacecraft over a certain time range

To demonstrate how these services would work one service has been added to the prototype

Middleware 1 And Ontological Layer

2.) Middleware routes message to appropriate service provider

Middleware 2 And Ontological Layer

3.) Awaiting SOAP daemon receives the message and calls the SSC interface software with the appropriate inputs

SSC Service residing on PC

This service provides access to the Satellite Situation Center (SSC) [2] and provides the ability to obtain position information for all spacecrafts covered by the SSC

Service has a dual purpose with second being the ability to demonstrate the heterogeneousness of the system and its transparency to the user. The service is run on a Windows PC while all other machines involved are Unix variants

# Design Benefits and Scaling Issues

Main Benefits

- Provides access to distributed data through one common interface
- Allows for searching and advanced querying of data as well as downloading
- Uses hostname and not IP address to find data providers.
- Uses SSL security for messaging
- Uses space science dictionary (SPASE+)
- Uses HTML for messaging which uses a port open to nearly every system

Scaling Issues

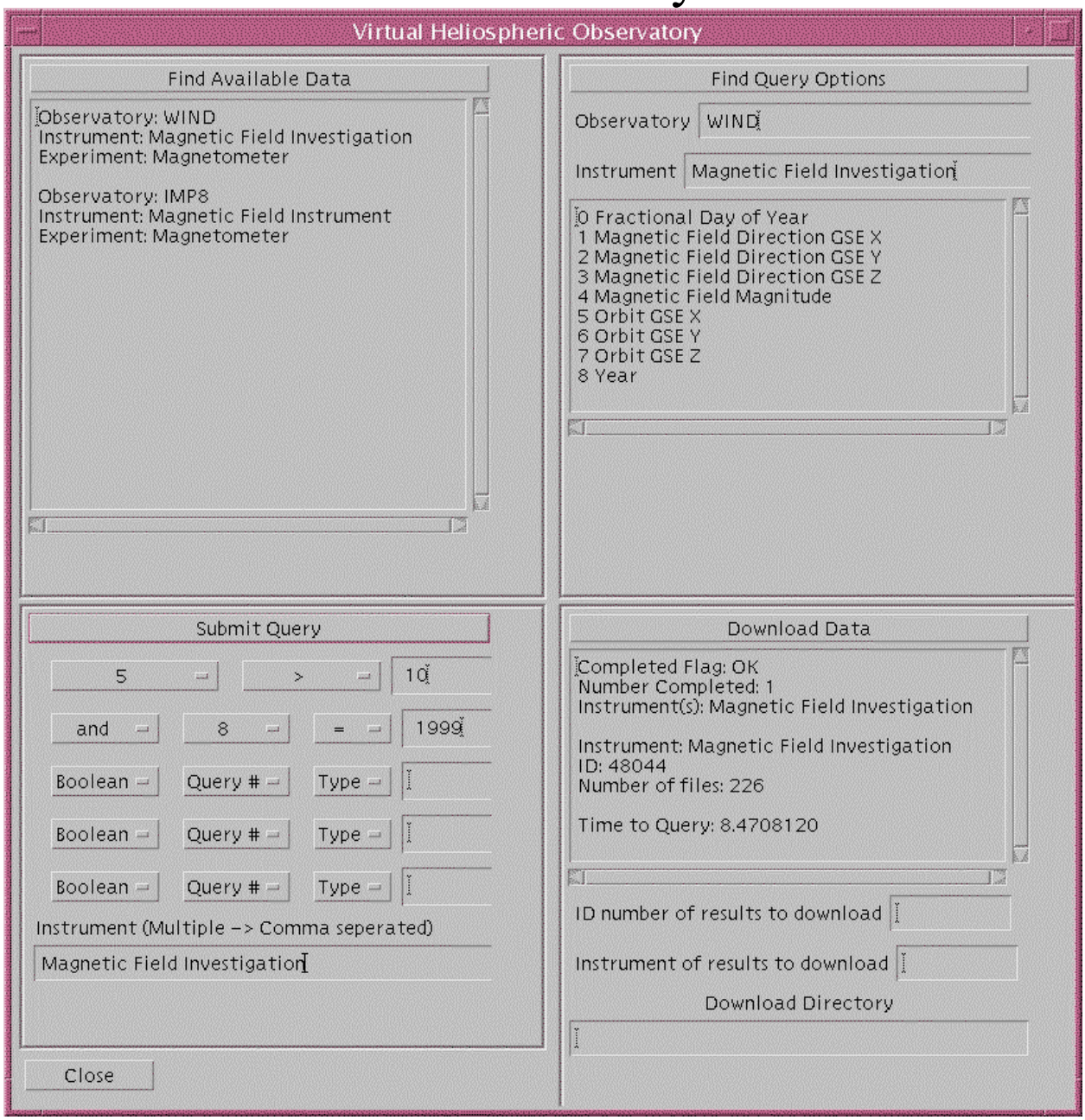
1. As number of users increases there is a need for more middleware nodes
2. Need for synchronization as data providers add backup and mirror sites
3. As data providers increase a peer-to-peer architecture will be tested to study data discovery and messaging for any possible benefits

# SPASE

- System provides a mapping between data provider naming and SPASE+ dictionary
- SPASE in still an ongoing and evolving effort. This prototype uses a variation of the dictionary known as SPASE+ that was available at the time of development.
- Allows users to see one naming convention and hides data provider specific names when querying
- Upon query submission names are converted back to data provider specific names
- Future implementation of VHO will require data provider metadata to be SPASE+ compliant

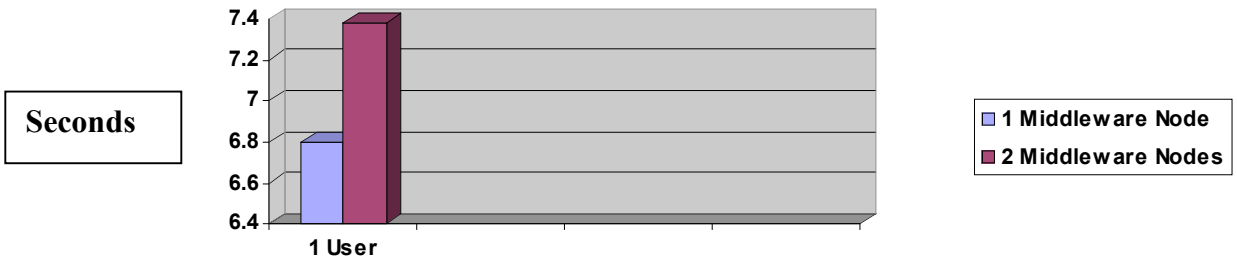
# Graphical User Interface to VHO

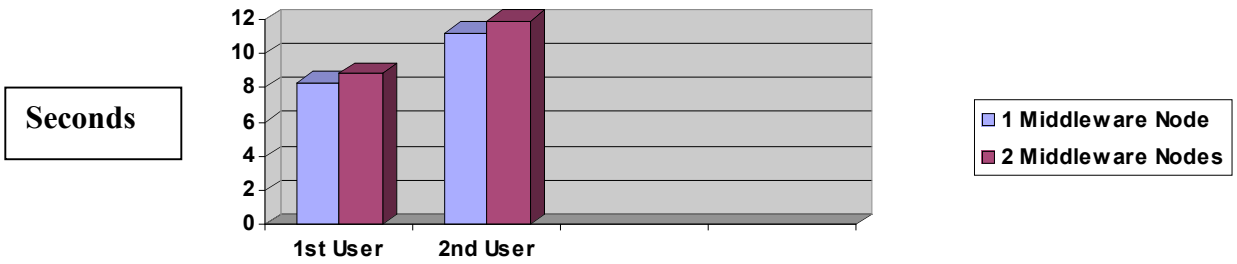
- Written in Interactive Data Language with SOAP access supplied through Perl code
- Shown with query submitted for WIND data containing GSE X position greater than 10 Earth Radii in the year 1999

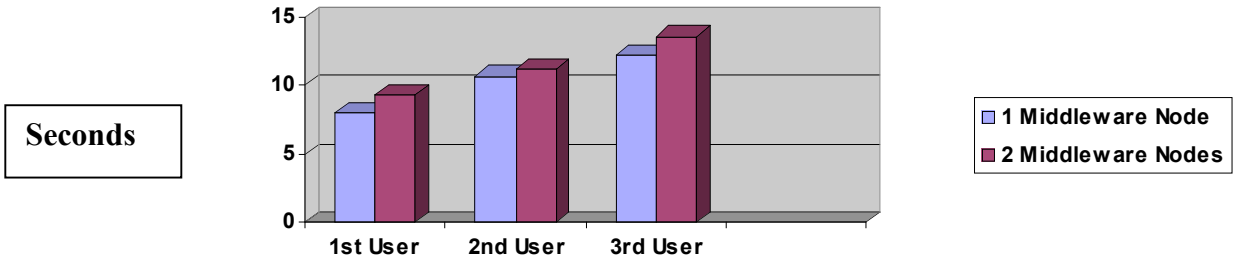


# Comparison of Query Times with Single Middleware Node and Two Middleware Nodes

- Two middleware nodes remove single point of failure
- Provides load balancing for times of large traffic volume
- Increase in time due to load checking is negligible



# Sample RDF File

- RDF Global Metadata Stored at Middleware

```
<rdf:RDF
    xmlns:wind_magnetometer="http://lepmfi.gsfc.nasa.gov/WIND#"
    xmlns:rdf ="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
    <rdf:Description about="WIND Magnetic Field Investigation">
        <wind_magnetometer:Observatory value="WIND" />
        <wind_magnetometer:Experiment_Type value="Magnetometer"
/>
        <wind_magnetometer:Instrument_Name value="Magnetic Field
            Investigation" />
        <wind_magnetometer:Observed_Physical_Parameter
                attribute1="Magnetic Field Magnitude"
                attribute2="Magnetic Field Direction GSE X"
                attribute3="Magnetic Field Direction GSE Y"
                attribute4="Magnetic Field Direction GSE Z"
                attribute5="Orbit GSE X"
                attribute6="Orbit GSE Y"
                attribute7="Orbit GSE Z"
                attribute8="Year"
                attribute9=”Fractional Day of Year” />
        <wind_magnetometer:Quantity_Name
                attribute1="F1"
                attribute2="BxGSE"
                attribute3="ByGSE"
                attribute4="BzGSE"
                attribute5="xGSE"
                attribute6="yGSE"
                attribute7="zGSE"
                attribute8=”Year”
                attribute9="Time" />
    </rdf:Description>
</rdf:RDF>
```

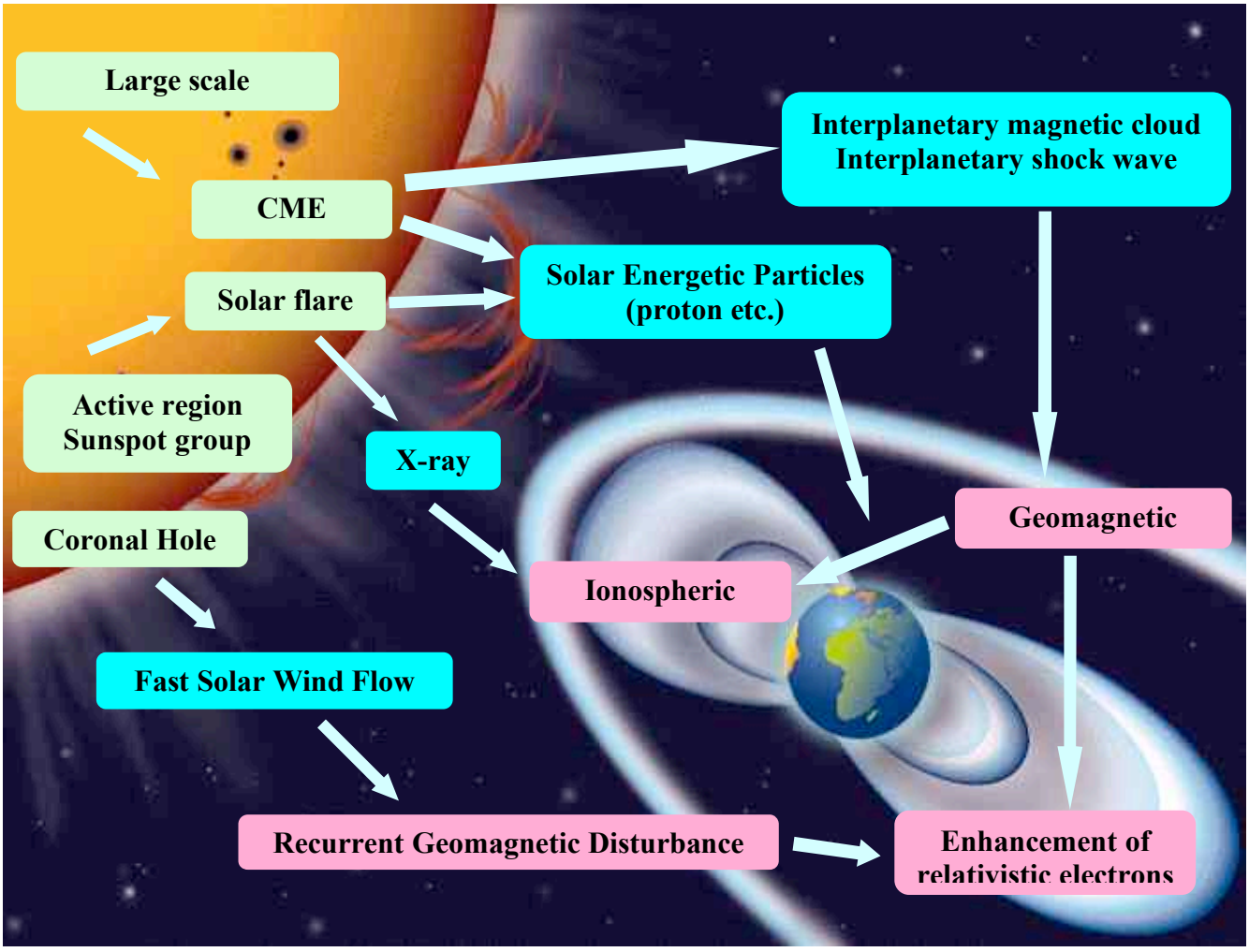
# Sample SOAP Query Message

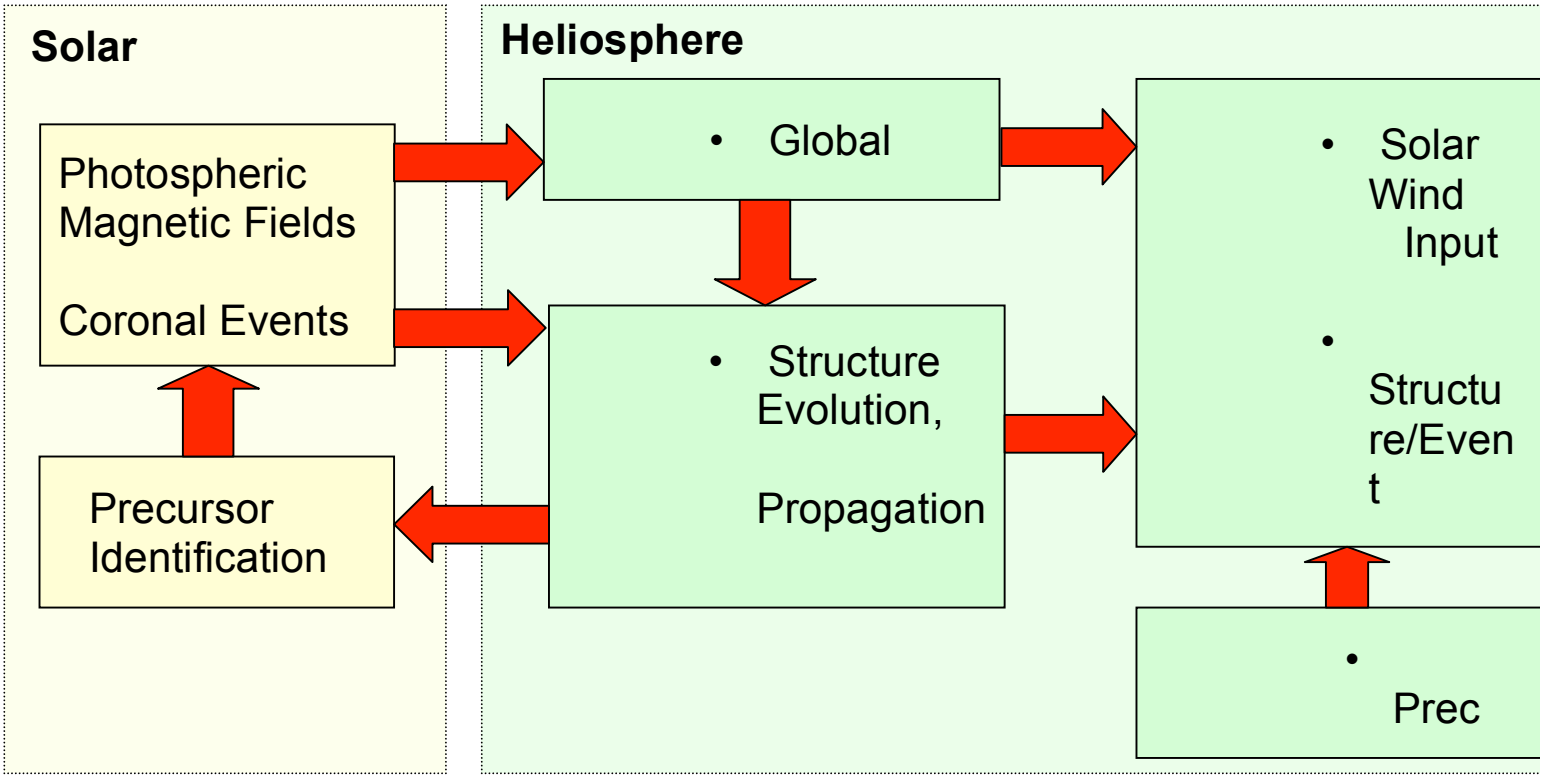
- SOAP message used in query

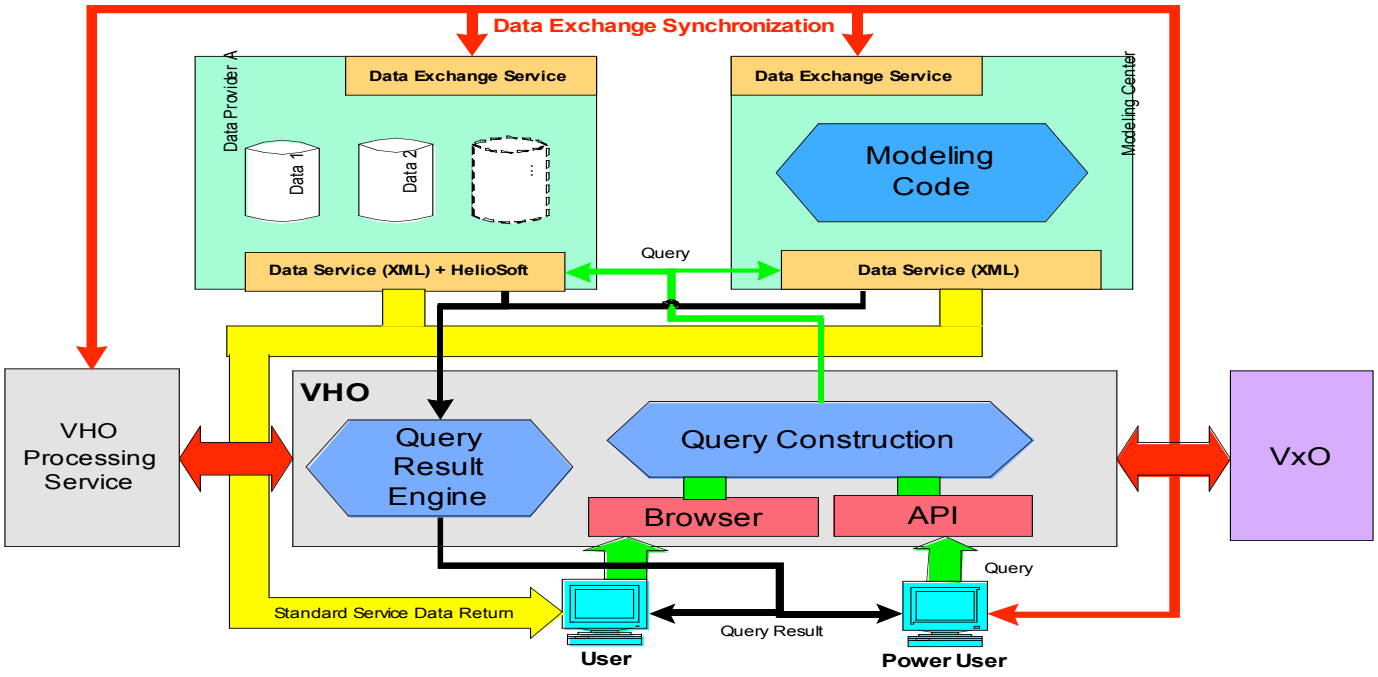
```
<?xml version="1.0" encoding="UTF-8"?>
<SOAP-ENV:Envelope
  xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"
  xmlns:xsi="http://www.w3.org/1999/XMLSchema-instance"
  xmlns:xsd="http://www.w3.org/1999/XMLSchema">
<SOAP-ENV:Header>
</SOAP-ENV:Header>
<SOAP-ENV:Body>
  <namesp1:query xmlns:namesp1="urn:QueryHandler">
    <providernum xsi:type="xsd:int">1</providernum>
    <provider xsi:type=”xsd:string”>WIND</provider>
    <numqueries xsi:type=”xsd:int”>1</numqueries>
    <query xsi:type=”xsd:string”>Orbit GSE X>10</query>
  </namesp1:query>
</SOAP-ENV:Body>
</SOAP-ENV:Envelope>
```

# Primary Objective of Virtual Observatories

**Facilitate data exchange and provide added value services for distributed data spanning all Sun Earth Connection (SEC) disciplines**






**Role**
**obtai**
**Obje**

# Availability

- Following this AGU meeting user software will be made available through the web site: http://vho.gsfc.nasa.gov

- We invite the community to test the software and submit any comments using the online form at the above mentioned web site

- Also following this AGU meeting will be the ability to submit queries directly from the web site.

# Acknowledgements


- The authors would like to thank Chris Howard for the development of the original SSCWeb SOAP interface

# References

[1] Szabo, Adam. “VHO Presentation to the LWS MOWG”, NASA HQ – June 19-20, 2003.

[2] SPASE, NSSDC News Article
http://nssdc.gsfc.nasa.gov/nssdc_news/mar03/dictionary.html

[3] Satellite Situation Center (SSCWeb) Operated by NASA/GSFC Space Physics Data Facility (SPDF) and the National Space Science Data Center (NSSDC)
http://sscweb.gsfc.nasa.gov